
Stephan Matzka

Künstliche Intelligenz in den Ingenieurwissenschaften

Maschinelles Lernen verstehen
und bewerten

 Springer Vieweg

Inhaltsverzeichnis

1 Grundlagen	1
1.1 Terminologie	2
1.1.1 Künstliche Intelligenz	3
1.1.2 Expertensysteme	4
1.1.3 Maschinelles Lernen	4
1.1.4 Künstliche neuronale Netze und Deep Learning	5
1.1.5 Big Data	7
1.1.6 IoT und Industrie 4.0	9
1.1.7 Zustandsüberwachung	9
1.1.8 Prädiktive Instandhaltung	10
1.2 Verfahren des maschinellen Lernens	10
1.2.1 Überwachtes Lernen	11
1.2.2 Unüberwachtes Lernen	13
1.3 Projektphasen	15
1.3.1 Klärung der Projektziele und -ressourcen	16
1.3.2 Daten visualisieren und analysieren	17
1.3.3 Daten vorbereiten und bereinigen	17
1.3.4 Maschinelle Modellbildung	18
1.3.5 Evaluation des maschinellen Modells	18
1.3.6 Betrieb und Überwachung	18
1.4 Verständnisfragen zu Kapitel 1	19
2 Daten visualisieren und analysieren	21
2.1 Tabellen	22
2.2 Lage	26
2.2.1 Mittelwert	26
2.2.2 Median	28
2.2.3 Modus	29
2.3 Streuung	30

2.3.1	Histogramm	30
2.3.2	Standardabweichung und Varianz	33
2.3.3	Quantile	36
2.3.4	Box-Plot	37
2.4	Statistische Zusammenhänge	40
2.4.1	Streudiagramm	41
2.4.2	Korrelation	42
2.4.3	Kovarianz	44
2.4.4	Korrelationskoeffizient	46
2.4.5	Bestimmtheitsmaß	49
2.4.6	Grand Tour	50
2.4.7	Korrelationsmatrix	51
2.4.8	Kreuztabelle	53
2.5	Zielsetzung der explorativen Analyse	55
2.6	Verständnisfragen und Übungen zu Kapitel 2	56
3	Daten vorbereiten und bereinigen	59
3.1	Erkennen von Problemen	59
3.1.1	Ort der Fehlererkennung	60
3.1.2	Methoden der Problemerkennung	61
3.2	Problemarten	62
3.2.1	Fehlende Werte	62
3.2.2	Ungültige Werte	62
3.2.3	Vermischte Werte	63
3.2.4	Werte in falschen Eigenschaften	63
3.2.5	Unplausible Einträge	64
3.2.6	Widersprüchliche Einträge	64
3.2.7	Duplikate	65
3.2.8	Abweichende Schreibweisen	65
3.3	Problembhebung	66
3.3.1	Manuelle Korrektur fehlerhafter Einträge	66
3.3.2	Statistische Korrektur fehlerhafter Einträge	67
3.3.3	Automatische Korrektur fehlerhafter Einträge	70
3.4	Harmonisierung von Zeitreihen	72
3.4.1	Definition einer gemeinsamen Zeitreihe	73
3.5	Auswahl und Kombination von Eigenschaften	77
3.5.1	Methoden zur Auswahl von Eigenschaften	78
3.5.2	Auswahl relevanter Eigenschaften	80

3.5.3	Erweiterung und Kombination von Eigenschaften	84
3.5.4	Hochdimensionale Eigenschaftsräume	85
3.6	Normalisierung von Werterräumen	93
3.7	Verständnisfragen und Übungen zu Kapitel 3	96
4	Überwachtes Lernen	99
4.1	k-nearest-Neighbor	101
4.1.1	Distanzmetriken	102
4.1.2	Klassifikationszeit	103
4.1.3	Anwendung zur Regression	104
4.2	Entscheidungsbaum	104
4.2.1	Trainingszeit	106
4.2.2	Qualitative Bewertung der Algorithmen	106
4.2.3	Anwendung zur Regression	107
4.3	Random-Forest	107
4.3.1	Generalisierung	107
4.3.2	Ensemble-Klassifikatoren	108
4.3.3	Anwendung zur Regression	110
4.3.4	Qualitative Bewertung der Algorithmen	110
4.4	Diskriminanzanalyse	111
4.4.1	Anwendung zur Regression	113
4.5	Künstliche neuronale Netze	114
4.5.1	Feed-Forward-Netz	117
4.5.2	Optimierung der Modellparameter	121
4.5.3	Aktivierungsfunktion	122
4.5.4	Rückpropagierung	124
4.5.5	Gradientenverfahren	126
4.5.6	Rekurrente Netzwerke	128
4.5.7	Anwendung zur Regression	129
4.5.8	Auslegung künstlicher neuronaler Netze	130
4.6	Support Vector Machines (SVM)	133
4.6.1	Maximierung des Trennabstandes	134
4.6.2	Nutzung von Kernen	134
4.6.3	Qualitative Bewertung der Algorithmen	136
4.6.4	Anwendung zur Regression	136
4.7	Gaußprozessregression	137
4.7.1	Vergleich zu Polynomen hoher Ordnung	138
4.7.2	Multiple Gaußprozessregression	139

4.8	Evaluation von Klassifikatoren	139
4.8.1	Grenzwertoptimierungskurve (ROC)	141
4.8.2	Wahrheitsmatrix	143
4.8.3	Bestimmung des unternehmerischen Nutzens	145
4.8.4	Kennzahlen für Klassifikatoren	147
4.8.5	Generalisierung	151
4.9	Evaluation von Regressionsmodellen	156
4.9.1	Bestimmtheitsmaß	157
4.9.2	Abweichung	157
4.9.3	Bestimmung des unternehmerischen Nutzens	160
4.9.4	Varianzanalyse	162
4.10	Verständnisfragen und Übungen zu Kapitel 4	166
5	Unüberwachtes Lernen	171
5.1	Zentrumsbasierte Verfahren	172
5.1.1	k-Means	172
5.1.2	k-Means++	175
5.1.3	Fuzzy c-Means Clustering	176
5.1.4	Vor- und Nachteile von zentrumsbasierten Verfahren	177
5.2	Dichtebasierte Verfahren	178
5.2.1	DBSCAN	178
5.2.2	Vor- und Nachteile von dichtebasierten Verfahren	179
5.3	Hierarchische Verfahren	180
5.3.1	Agglomeratives Clustering	181
5.3.2	Divisives Clustering	184
5.4	Evaluation von Clustering-Ergebnissen	186
5.4.1	Manuelle Beurteilung durch Expert:innen	187
5.4.2	Silhouetten-Diagramm	187
5.4.3	Dendrogramme	192
5.5	Verständnisfragen und Übungen zu Kapitel 5	195
6	Musterlösungen zu den Verständnisfragen und Übungen	199
	Literaturverzeichnis	211
	Stichwortverzeichnis	217